



Published in final edited form as:

*Epidemiology*. 2015 September ; 26(5): 727–732. doi:10.1097/EDE.0000000000000353.

## Negative Control Outcomes and the Analysis of Standardized Mortality Ratios

DB Richardson<sup>1</sup>, A Keil<sup>1</sup>, Tchetgen E Tchetgen<sup>2</sup>, and GS Cooper<sup>3</sup>

<sup>1</sup>Department of Epidemiology, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA 27599

<sup>2</sup>Department of Biostatistics, Harvard University, Boston, MA USA 02115

<sup>3</sup>National Center for Environmental Assessment, US Environmental Protection Agency, Washington, DC, USA 20460

### Abstract

In occupational cohort mortality studies, epidemiologists often compare the observed number of deaths in the cohort to the expected number obtained by multiplying person-time accrued in the study cohort by the mortality rate in an external reference population. Interpretation of the result may be difficult due to non-comparability of the occupational cohort and reference population. We describe an approach to estimate an adjusted standardized mortality ratio (aSMR) to control for bias due to unmeasured differences between the occupational cohort and the reference population. The approach draws on methods developed for the use of negative control outcomes. Conditions necessary for unbiased estimation are described, as well as looser conditions necessary for bias reduction. The approach is illustrated using data on bladder cancer mortality among male Oak Ridge National Laboratory workers. The SMR for bladder cancer was elevated among hourly-paid males (SMR=1.90; 1.27, 2.72) but not among monthly-paid males (SMR=0.96; 0.67, 1.33). After indirect adjustment using the proposed approach, the mortality ratios were similar in magnitude among hourly- and monthly-paid men (aSMR=2.22; 1.52, 3.24; and, aSMR=1.99; 1.43, 2.76, respectively). The proposed adjusted SMR offers a complement to typical standardized mortality ratio analyses.

### Keywords

cohort studies; mortality study; occupational diseases

Evaluations of potential carcinogens, such as those conducted by the International Agency for Research on Cancer and the National Toxicology Program, play an important role in occupational and environmental protection. For an agent to be classified as a known carcinogen there must be evidence from studies of human populations; often, such epidemiological evidence derives from occupational cohort mortality studies.

One of the commonly-used measures of relative mortality in occupational cohort studies is the ratio of observed to expected deaths, the latter obtained by multiplying person-time accrued in an occupational cohort by the mortality rate in an external reference population, usually all residents of a nation or region. When the expected number is computed by taking into account some covariates by indirect standardization, the observed-to-expected ratio is called a standardized mortality ratio (SMR).

Assuming that the reference population mortality rates accurately represent the mortality rates that would have been observed if the occupational cohort was not exposed to the potential carcinogen of interest, the SMR quantifies the effect of the potential carcinogen on mortality rates. If the assumption does not hold then the SMR may yield a biased estimate of this effect measure. This potential for bias poses an important obstacle to the use of SMR analyses in the evaluation of an agent's role as a human carcinogen. An SMR of unity could reflect absence of an exposure effect, or it could reflect bias that is masking the exposure's effect. Judgments regarding the direction and magnitude of bias in SMRs therefore play a role in interpreting this type of evidence when used for such evaluations. The ubiquity of SMRs below unity for major categories of cause of death in occupational cohort studies, often referred to as "the healthy worker effect" has led some authors to advocate for abandoning SMR analyses altogether.<sup>1</sup>

We describe an approach to estimate an adjusted standardized mortality ratio (aSMR) to reduce bias in SMR analyses. The approach draws on methods developed for the use of negative control outcomes.<sup>2</sup> The purpose of the negative control is to reproduce conditions that cannot involve the causal effect of exposure but do involve the same sources of bias that affect the association of primary interest. Conditions necessary for unbiased estimation are described, as well as looser conditions necessary for reduction of bias in the adjusted estimate relative to the standard SMR.

## METHODS

The setting of interest is an evaluation in which occupational cohort mortality data are used to assess whether an agent is a human carcinogen. Suppose we stratify the study cohort into  $k=1 \dots K$  subgroups based on levels of confounders (e.g., five-year categories of age), where  $I_{1k}$  is the observed rate of death due to the outcome of interest in the cohort in stratum  $k$ , and  $I_{0k}$  is the (counterfactual) rate of death due to the outcome of interest that would have been observed had the cohort not been exposed to the occupational carcinogen of interest.

Within each stratum  $k$ , we wish to compare the rate of death due to the outcome of interest to the rate that would have been observed if the occupational cohort had not been exposed to the carcinogen of interest. A simple comparative statistic, for each stratum, is the rate ratio:

$$\frac{I_{1k}}{I_{0k}} = \exp(\alpha_k), \text{ for } k=1 \dots K, \quad (\text{Equation 1})$$

with  $\alpha_k$  denoting the log of the stratum-specific rate ratios, as when estimated in a log-linear regression. The parameters,  $\alpha_k$ , are the target parameters of primary interest that we would like to estimate.

Unfortunately, we do not get to see the counterfactual rates,  $I_{0k}$ . Instead epidemiologists often calculate comparative statistics using stratum-specific external reference rates,  $I_{Rk}$ , that may differ from the counterfactual rates. We can denote this deviation by  $\delta_k$ , using the expression

$$I_{Rk} = I_{0K} \exp(\delta_k), \text{ for } k=1 \dots K.$$

Comparing the observed stratum-specific rates in the cohort to the reference population rates yields,

$$\frac{I_{1k}}{I_{Rk}} = \frac{I_{1k}}{I_{0k} \exp(\delta_k)} = \exp(\alpha_k - \delta_k), \text{ for } k=1 \dots K. \quad (\text{Equation 2})$$

We might combine these stratum-specific rate ratios into a single summary figure; a weighted mean of the stratum-specific rate ratios can be obtained, where the weights are chosen to minimize the standard error of the weighted mean (Appendix). Usually an SMR is calculated for such data; if this is done using the usual formula then a numerically-equivalent summary measure is obtained.<sup>3</sup> This is because the approach in the Appendix for calculating a weighted mean of the stratum-specific rate ratios is simply an alternative to the usual formula for calculating an SMR.<sup>4,5</sup>

If the reference population mortality rates accurately represent the mortality rates that would have been observed had the occupational cohort been unexposed (i.e.,  $\delta_k=0$ ) then a summary SMR based on the external reference rates summarizes the stratum-specific causal rate ratio (Equation 1).<sup>6</sup> However, the ubiquity of SMRs below unity for major categories of cause of death in occupational cohort studies, often referred to as “the healthy worker effect”, suggests a common problem of non-comparability of external reference rates to counterfactual rates.

## Negative Control Outcome

How can we adjust the rate ratios described by the expression in Equation 2 to better estimate the contrasts of interest (Equation 1)? One way is by leveraging assumptions external to the study data about a negative control outcome. The purpose of the negative control is to reproduce a condition that arguably cannot involve the causal effect of exposure but does involve the same sources of bias (confounding or selection) that affect the association of primary interest.<sup>2,7,8</sup> Figure 1 illustrates an ideal negative control outcome for our purposes. Occupational exposure is not a cause of the negative control outcome. There is an unmeasured factor, however, that is associated with occupational exposure, risk of death due to the outcome of interest, and risk of death due to the negative control outcome.

Suppose  $J_{1k}$  are rates of the negative control outcome in the occupational cohort, and  $J_{0k}$  are expected rates of the negative control in the absence of exposure. Again, stratum-specific external reference rates for the negative control,  $J_{Rk}$ , may differ from the expected rates for the negative control outcome in the absence of exposure; this difference can be described by the parameters,  $\epsilon_k$ , under the model:  $J_{Rk} = J_{0k} \exp(\epsilon_k)$ . An expression for the comparative

statistic for the rate of the negative control outcome in the occupational cohort to the stratum-specific external reference rate for the negative control is:

$$\frac{J_{1k}}{J_{Rk}} = \frac{J_{1k}}{J_{0k} \exp(\varepsilon_k)} = \exp(-\varepsilon_k), \text{ for } k=1 \dots K, \quad (\text{Equation 3})$$

since the rate of the negative control outcome is not affected by the exposure of interest.

That is, we are assuming that our choice of negative control outcome satisfies  $\frac{J_{1k}}{J_{0k}}=1$ .

### Indirect Adjustment

Complete adjustment for confounding is possible if there is equivalence of bias magnitude for the negative control outcome ( $\varepsilon_k$ ) and outcome of primary interest ( $\delta_k$ ). Using the negative control outcome, we can derive an adjusted comparative statistic for each stratum:

$$\frac{\frac{I_{1k}}{I_{Rk}}}{\frac{J_{1k}}{J_{Rk}}} = \exp(\alpha_k + \varepsilon_k - \delta_k), \text{ for } k=1 \dots K. \quad (\text{Equation 4})$$

By calculating the weighted mean of the stratum-specific comparative statistics, where the weights are chosen to minimize the standard error of the weighted mean, a summary figure can be obtained. We refer to this summary figure as an adjusted SMR (aSMR).

Bias is reduced, though not entirely eliminated, as long as  $|\varepsilon_k - \delta_k| < |\delta_k|$ . This condition holds, for example, when  $0 < \delta_k$ , as long as  $\varepsilon_k$  falls within the range  $0 < \varepsilon_k < 2\delta_k$ , which implies that the ratio of external reference rates for the negative control and outcome of interest to counterfactual rates do not differ by more than a factor of  $\exp(2) = 7.4$ . Therefore, over a wide range of conditions, the aSMR (derived from Equation 4) will yield a less biased estimate of the quantity of interest (Equation 1) than the traditional SMR (derived from Equation 2).

The appendix provides SAS code for estimation of the aSMR and associated confidence intervals and can be applied to data derived from a life table program that is freely available.<sup>9,10</sup> Table 1 lists the assumptions discussed above that are necessary for the aSMR to reduce bias.

**Example**—A cohort of 16,912 male Oak Ridge National Laboratory (ORNL) workers who were hired prior to 1985 and who worked at least 30 days, with complete information on name, social security number, date of birth, and date of first hire was assembled. Vital status through December 31, 2008 was ascertained through searches of Social Security Administration records and the National Death Index (NDI). We used the NDI-Plus service to obtain underlying cause of death for deceased workers identified by the NDI. For deaths prior to 1979, cause of death information was coded according to the Eighth revision of the International Classification of Diseases (ICD); for deaths occurring in 1979 and later, cause of death information was coded to the ICD revision in effect at the time of death. If there was no death indication for a worker and they were confirmed to be alive on January 1, 1979

or later by the Social Security Administration or by ORNL's employment records then they were assumed to be alive as of December 31, 2008. Those lost to follow-up before January 1, 1979 were only considered alive until the date last observed. The mortality experience of the cohort was analyzed using the life table analysis system (LTAS)<sup>9,11</sup>. SMRs and aSMRs were compared, the latter estimated by modeling the observed number of deaths in strata defined by five-year categories of age and calendar period, sex, and race (white or non-white). These analyses focus on deaths due to bladder cancer, where the occupational exposure of interest is ionizing radiation, and ischemic heart disease is taken as the negative control outcome for all calculation of aSMRs. Analyses were conducted for subgroups defined by white-collar (monthly-paid) and blue-collar (hourly-paid) men.

## RESULTS

There were 101 deaths due to bladder cancer. The SMR for bladder cancer was elevated among hourly-paid males (SMR=1.90; 1.27, 2.72) but not among monthly-paid males (SMR=0.96; 0.67, 1.33). After indirect adjustment (Table 2), the mortality ratios were similar in magnitude among hourly- and monthly-paid workers (aSMR=2.22; 1.52, 3.24; and, aSMR=1.99; 1.43, 2.76, respectively). The heterogeneity in SMR appears to be due to paycode differences in comparability of occupational cohort to reference rates, and this heterogeneity is reduced by the proposed indirect adjustment approach.

## DISCUSSION

The illustrative analysis of mortality among ORNL workers shows how minimizing "healthy worker" effects reduced evidence of apparent heterogeneity in bladder cancer SMRs between hourly- and monthly-paid ORNL workers. A naïve interpretation of the bladder cancer SMRs for hourly-paid (SMR=1.90) and monthly-paid (SMR=0.96) men might lead an investigator to conclude that this pattern reflects higher occupational exposure to bladder carcinogens among blue-collar than white-collar workers at facility. However, prior research on carcinogenic exposures (e.g., ionizing radiation) at ORNL did not suggest that white collar workers had substantially less exposure than blue collar workers. An alternative explanation is that the external reference rates are a better proxy for the counterfactual bladder cancer rates that would be observed for blue-collar workers than they are for the white-collar workers. The latter explanation is reasonable because white-collar workers at ORNL tended to be highly educated technical professionals who exhibited substantial deficits in mortality for a range of other smoking-related causes of death.

In an analysis of aSMRs there was little evidence of heterogeneity in bladder cancer observed-to-expected mortality ratios between hourly- and monthly-paid workers. The finding is supportive of the conclusion that the difference in bladder cancer SMRs by pay code was an artifact of bias due to non-comparability of the counterfactual reference rates for white collar workers and the external reference population; such conclusions hold if one accepts that the conditions for the aSMR to yield less biased results appear reasonable in this example (Table 1).

Interpretation of the traditional SMR is challenging because the occupational cohort and reference population may differ (within strata of confounders, such as age and calendar period) with respect to factors other than the exposure of interest. This is a failure of the conditional exchangeability assumption.<sup>12</sup> The proposed aSMR offers a potentially useful complement to the classical SMR that may reduce confounding bias through indirect adjustment using a negative control outcome.

Under the ideal case of bias equivalence, there is complete elimination of bias in the adjusted SMR. However, failing the ideal case, under a wide range of conditions the adjusted SMR will be less biased than the standard SMR. Bias reduction occurs if  $\varepsilon_k$  and  $\delta_k$  have the same sign, and  $|\varepsilon_k|$  lies between zero and twice  $|\delta_k|$ . While the sign and magnitude of  $\varepsilon_k$  can be determined from the negative control outcome,  $\delta_k$  is unknown. However, in settings where a healthy worker bias is expected, for example,  $\delta_k$  might be considered positive. While these conditions are not testable assumptions, they would be supported if there is belief that a moderate or strong healthy worker bias was operating, and  $\varepsilon_k$  was relatively small.

Under certain conditions we can relax the assumption of bias equivalence, yet still obtain complete control for confounding with this approach. If the relation between an unmeasured confounder ( $U$ ) and the negative control outcome, and that between  $U$  and the potential outcome for the disease of interest in the absence of exposure ( $Y_0$ ) are monotone at the individual level, then bias is eliminated entirely, even if the association between  $U$  and  $N$  is quite distinct from that between  $U$  and  $Y$ .<sup>13</sup>

Interestingly, Equation 4 can be equivalently expressed without reference to the observed person-time in the occupational cohort. This suggests an appealing aspect of the aSMR. Unlike the traditional SMR, the aSMR can be estimated in settings in which enumeration of person-time at risk is infeasible. For example, some occupational mortality studies draw upon a registry of events (deaths or disease) but do not have access to information necessary to calculate person-time at risk.<sup>14</sup> The aSMR may be calculated as an alternative to the proportionate mortality ratio, which is often used in such settings.

Furthermore, we may note that Equation 4 is algebraically equivalent to a stratum-specific mortality odds ratio.<sup>15</sup> Previous papers on mortality odds ratios framed the effect measure in terms of a cumulative case-control study design: cases represented events ascertained over a follow-up period and controls are selected from a set of reference causes of death.<sup>16,17</sup> In contrast, Equation 4 is expressed in terms of estimation of an underlying rate ratio parameter for a specified exposure contrast, using a negative control outcome to reduce bias in the stratum-specific rate ratio. The current work provides a connection between earlier work on analysis of cohort data using a mortality odds ratio and contemporary work on the logic of analysis using negative control outcomes. In the previous literature on the mortality odds ratio, the choice of auxiliary cause of death was framed as the problem of identifying a set of causes of death for which exposure is not a risk factor (for mortality proportions). Extending this, we show that beyond using the negative control outcome as a reference outcome, it can be used for bias reduction. This becomes the basis for an approach to reduce a major

limitation of SMR analysis: the “healthy worker effect.” Of course, a plausible negative control outcome that meets the assumptions may not be available in many settings.

We have framed the causal contrast of interest in terms of a ratio of the observed rate of an outcome of interest to the counterfactual rate of that outcome in the absence of exposure. The SMR is often discussed as the ratio of observed to expected deaths (rather than rates). These are equivalent assuming that exposure does not affect the distribution of person-time.

Interpretation of the traditional SMR requires one set of unverifiable assumptions (the reference rates represent the rate that would be seen in the cohort in the absence of exposure). Interpretation of the proposed aSMR requires a different set of unverifiable assumptions: the negative control outcome is not caused by the occupational exposure, but is impacted by similar bias factors (Table 1). While each approach requires unverifiable assumptions, the proposed aSMR may serve as a useful complement to traditional SMRs; in some cases, the opportunity to assess results under different assumptions regarding confounding may help investigators to better triangulate estimation of the true causal hazards ratio of interest.

## REFERENCES

1. Geiger, HJ.; Rush, D.; Michaels, D. Dead Reckoning: A critical review of the Department of Energy's Epidemiologic Research. Washington, DC: Physicians for Social Responsibility; 1992.
2. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: A tool for detecting confounding and bias in observational studies. *Epidemiology*. 2010; 21(3):383–388. [PubMed: 20335814]
3. Checkoway, H.; Pearce, N.; Kriebel, D. Monographs in Epidemiology and Biostatistics. Second. Oxford: Oxford University Press; 2004. Research Methods in Occupational Epidemiology.
4. Breslow, NE.; Day, NE. Statistical Methods in Cancer Research: The Design and Analysis of Cohort Studies. Vol. II. Lyon: International Agency for Research on Cancer; 1987.
5. McNamee R. Regression modelling and other methods to control confounding. *Occup Environ Med*. 2005; 62(7):500–506. 472. [PubMed: 15961628]
6. Maldonado G, Greenland S. Estimating causal effects. *Int J Epidemiol*. 2002; 31(2):422–429. [PubMed: 11980807]
7. Tchetgen, Tchetgen E. The control outcome calibration approach for causal inference with unobserved confounding. *Am J Epidemiol*. 2014; 179(5):633–640. [PubMed: 24363326]
8. Richardson DB, Laurier D, Schubauer-Berigan MK, Tchetgen ET, Cole SR. Assessment and indirect adjustment for confounding by smoking in cohort studies using relative hazards models. *Am J Epidemiol*. 2014; 180(9):933–940. [PubMed: 25245043]
9. Robinson CF, Schnorr TM, Cassinelli RT 2nd, Calvert GM, Steenland NK, Gersic CM, Schubauer-Berigan MK. Tenth revision U.S. mortality rates for use with the NIOSH Life Table Analysis System. *J Occup Environ Med*. 2006; 48(7):662–667. [PubMed: 16832222]
10. Schubauer-Berigan MK, Hein MJ, Raudabaugh WM, Ruder AM, Silver SR, Spaeth S, Steenland K, Petersen MR, Waters KM. Update of the NIOSH life table analysis system: a person-years analysis program for the windows computing environment. *Am J Ind Med*. 2011; 54(12):915–924. [PubMed: 22068723]
11. Steenland K, Beaumont J, Spaeth S, Brown D, Okun A, Jurcenko L, Ryan B, Phillips S, Roscoe R, Stayner L, Morris J. New developments in the Life Table Analysis System of the National Institute for Occupational Safety and Health. *Journal of Occupational Medicine*. 1990; 32(11):1091–1098. [PubMed: 2258764]
12. Cole SR, Hernan MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*. 2008; 168(6):656–664. [PubMed: 18682488]



13. Sofer, T.; Cole, SR.; Richardson, DB.; Tchetgen, Tchetgen E. On Relations Between Difference-in-Difference and Negative Control Outcomes: Identifying Assumptions and Some Generalizations. Boston, MA: 2014.
14. Clapp RW. Mortality among US employees of a large computer manufacturing company: 1969–2001. *Environ Health*. 2006; 5:30. [PubMed: 17052328]
15. Stewart W, Hunting K. Mortality odds ratio, proportionate mortality ratio, and healthy worker effect. *Am J Ind Med*. 1988; 14(3):345–353. [PubMed: 3189351]
16. Wang JD, Miettinen OS. The mortality odds ratio (MOR) in occupational mortality studies--selection of reference occupation(s) and reference cause(s) of death. *Ann Acad Med Singapore*. 1984; 13(2 Suppl):312–316. [PubMed: 6497331]
17. Miettinen OS, Wang JD. An alternative to the proportionate mortality ratio. *Am J Epidemiol*. 1981; 114(1):144–148. [PubMed: 7246522]

## APPENDIX

A simple tabular example is provided to illustrate the data structure and SAS code that may be used to implement this approach.

The data in Table A1 were generated under a model where the true stratum-specific rate ratios for the outcome of interest equal two (i.e.,  $I_{1k}/I_{0k} = 2$ ) and the stratum-specific rate ratios for the negative control outcome equal unity (i.e.,  $J_{1k}/J_{0k} = 1$ ). Stratum-specific external reference rates differ from counterfactual rates,  $I_{Rk} = I_{0k} \exp(\delta_k)$  and  $J_{Rk} = J_{0k} \exp(\epsilon_k)$ , where  $\delta_k = \epsilon_k = 0$ . The data in Table A1 consist of person-time and events for the outcome of interest, a negative control outcome, and external reference rates for the outcome of interest and the negative control outcome, where  $T_{1k}$  is the number of person-years in the occupational cohort in stratum  $k$ ,  $Y_{1k}$  is the number of deaths due to the cause of interest in the cohort in stratum,  $k$ .

The four stratum specific rate ratios are close in value and therefore it seems reasonable to combine them into a summary value. The standardized mortality ratio (SMR) can be calculated, in the usual manner, as  $\sum Y_{1k} / \sum T_{1k} I_{Rk}$ . This is equivalent to the weighted average of the stratum-specific rate ratios,  $[Y_{1k}/T_{1k}]/I_{Rk}$ , where the weight for stratum  $k$  is  $T_{1k} I_{Rk} / \sum T_{1k} I_{Rk}$ .

The data in Table A1 could be assembled in a SAS data set and analyzed using the sample code provided in Figure A1. Using SAS PROC GENMOD, a Poisson regression model may be fitted to these data to estimate the SMR<sup>4</sup>, where the log of the product of the external reference rates and person-time serve as an offset (Figure A2).

## Adjusted SMR

The SMR=1.32; this is a biased estimate of the desired summary rate ratio ( $I_{1k}/I_{0k} = 2.0$ ) because  $\delta_k \neq 0$ . The manuscript proposes calculation of an adjusted SMR (aSMR) using a negative control outcome to reduce this form of bias. The aaSMR can be obtained by fitting a Poisson regression model where the log of the product of the number of negative control outcome events and the ratio of external reference rates for the outcome of interest and negative control outcome, serve as an offset (Figure A3). The aSMR (aSMR=2.00; 95% CI: 1.72, 2.32) equals the desired summary ratio of the observed to counterfactual rates ( $I_{1k} / I_{0k}$



$=2.0$ ; 95% CI: 1.72, 2.32) because the reference rates for the negative control outcome,  $J_R$ , differ from counterfactual reference rates  $J_0$  by a factor  $\varepsilon_k$  that equals  $\delta_k$ .

If there are strata with no negative control outcome events then calculation of this offset will be problematic because it involves taking the log of zero. This may be handled by modifying the last line of SAS code in Figure A1 as follows,  $\text{offset2}=\log(I\_R / J\_R * \max(0.001,N))$ ;

**Table A1**

Hypothetical cohort data consisting of person-time, deaths, and negative controls.

Age ( <i>k</i> )	Occupational cohort			Extl ref.	Summary statistics		Neg. control	
	Deaths ( $Y_{1k}$ )	P-yrs ( $\times 10^3$ ) ( $T_{1k}$ )	Death Rate ( $Y_{1k}/T_{1k}$ )	Ref rate ( $I_{Rk}$ )	Rate ratio ( $Y_{1k}/T_{1k}/I_{Rk}$ )	Weight <sup>*</sup>	Obs ( $N_k$ )	Ref rate ( $J_{Rk}$ )
55–59	6	1.200	5.000	3.375	1.48	0.030831	9	10.12
60–64	22	2.340	9.402	7.013	1.34	0.124924	14	8.92
65–69	98	3.750	26.133	20.493	1.28	0.584957	110	46.00
70–74	48	0.975	49.231	34.931	1.41	0.259287	60	87.33
Total	174	8.265			1.32 <sup>*</sup>	1.0	193	

\* Weighted average of stratum-specific rate ratios, where weight for stratum  $k$  is  $T_{1k} I_{Rk} / \sum T_{1k} I_{Rk}$ .

Figure A1. Sample code to assemble the data in Appendix Table A1 as a data set for analysis in the SAS statistical package.

```

data a;
input
k      Y      T      N      I_R      J_R ;
cards;
1      6      1.200      9      3.3746      10.12
2      22     2.340      14     7.0129      8.92
3      98     3.750     110    20.4926     46.00
4      48     .975      60    34.9309     87.33;

data a; set a;
offset1=log(T * I_R);
offset2=log(I_R / J_R * N);

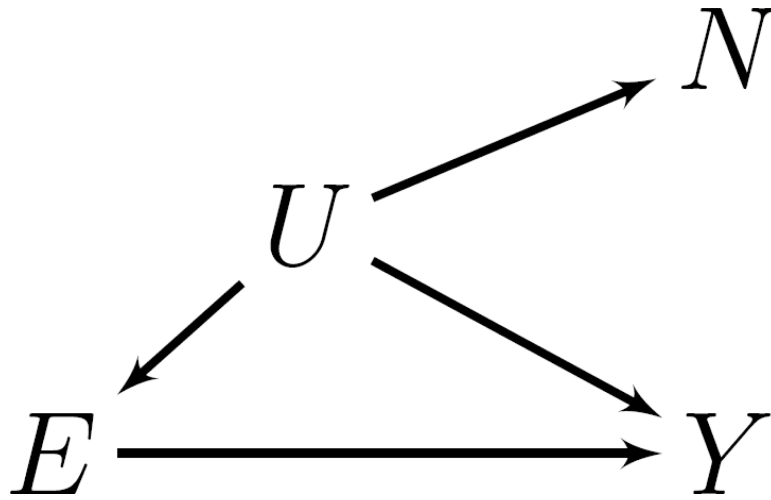
```

Figure A2. Illustrative SAS code to obtain a weighted summary of the stratum-specific rate ratios, that equals the standard SMR.

```
proc genmod data=a;  
model Y = / dist=poisson link=log offset=offset1;  
estimate 'SMR' int 1 / exp ; run;
```

Figure A3. Illustrative SAS code to obtain the adjusted SMR described in this paper.

```
proc genmod data=a;  
model Y = / dist=poisson link=log offset=offset2;  
estimate 'adjusted SMR' int 1 / exp ; run;
```



**Figure 1.**

Directed acyclic graph illustrating an ideal negative control outcome. For one stratum,  $k$ .  $E$  denotes exposure,  $Y$  denotes outcome of interest,  $N$  denotes negative control outcome, and  $U$  denotes unmeasured causes of  $E$ ,  $N$  and  $Y$ .

Table 1

Assumptions required for the aSMR to reduce bias.

i	The exposure of interest is not a cause of the negative control outcome.
ii	There is an open backdoor path between the exposure of interest and outcome of primary interest, as well as with the negative control outcome (see Figure 1).
iii	The direction of bias for the negative control outcome and outcome of primary interest is the same (i.e., $\varepsilon_k$ and $\delta_k$ have the same sign), and $ \varepsilon_k $ lies between zero and twice $ \delta_k $ .

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**

Traditional and adjusted standardized mortality ratios for bladder cancer. Men employed at Oak Ridge National Laboratory. Oak Ridge, Tennessee, 1943–2008.

	Blue collar (hourly-paid)	White collar (monthly-paid)
Traditional SMR (95%CI)	1.90 (1.27, 2.72)	0.96 (0.67, 1.33)
Adjusted SMR <sup>†</sup> (95%CI)	2.22 (1.52, 3.24)	1.99 (1.43, 2.76)

<sup>†</sup> Indirect adjustment using ischemic heart disease as negative control